DOI:
10.1063/1.4890343

Document Version
Peer reviewed version

Universal Tight Binding Model for Chemical Reactions in Solution and at Surfaces: II Analysis of Solvent Polarizability, Mixing and Diffusivity

A. Y. Lozovoi,[1] T. J. Sheppard,[1] D. L. Pashov,[2] J. J. Kohanoff,[1] and A. T. Paxton[2]
[1)] *Atomistic Simulation Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, U.K.*
[2)] *Department of Physics, King's College London, Strand, London WC2R 2LS, UK*

A new water model intended for use in condensed phase simulations in conjuction with the self consistent polarizable ion tight binding method is developed. The model is applied to water monomer, dimer, hexamers, ice, and liquid water where it demonstrates good agreement with theoretical results obtained by more accurate methods, such as DFT and CCSD(T), and with experiment. In particular, the temperature dependence of the dielectric constant and self diffusion coefficient in liquid water predicted by the model, closely reproduces experimental curves in the temperature interval between 260 K and 350 K.

INTRODUCTION

This is the second in a series of three papers describing a universal tight binding model for the electronic structure and interatomic forces in condensed phases. In the first, referred to as I[1] in what follows, we presented a scheme for molecular dynamics simulations of organic molecules and here we show how the method is extended for systems in polar solvents. In particular, we focus on water and later in the paper we discuss methanol and water–methanol mixtures.

Water is of course one of the most fascinating substances and, at the same time, a most important compound which ultimately permits the existence of life on Earth. Hence the fact that water has remained a focus of scientific attention for more than a century.

Theoretical simulations of water are done nowadays using a variety of methods ranging from density functional theory (DFT) to empirical force field models (SPC/E[2] and TIP4P[3] models and their derivatives being most popular during the last couple of decades). The tight binding (TB) approach is the method of choice if the required size of the system (or the length of the simulation in case of molecular dynamics) is outside the DFT range, but the effect of charge transfer or bond breaking, not captured in force field models, cannot be ignored.

Earlier it was demonstrated by some of us[4] that the self-consistent polarizable ion tight binding method[5,6] provides an excellent framework for the description of water. In this method, in addition to on site energies and hopping integrals, atomic species are assigned point multipoles up to a specified angular momentum. These point multipoles are not fixed but rather sought self-consistently.

Here we present a new TB water model to be used within the framework of the self-consistent polarizable ion method. The development of the new model was specifically tailored to its application to condensed phases of water, *i.e.* ice and liquid water. We also corrected a few pathological features that were discovered in the previous water model of Ref. 4 (see Secs. I and IV).

We expect the new water model to be suitable for simulations of organic molecules solvated in water, and also require that our model correctly describes an interface between water and metal oxides. In terms of model parameters, such "universality" means in particular that oxygen in water, oxygen in organic molecules, and oxygen in metal oxides must be the same entity. Hence, the development of the water model was linked to the development of the TB models for titania and CHO hydrocarbons. The two latter models are presented in detail elsewhere.[1,7]

In I we gave a brief summary of the self consistent polarizable ion tight binding (PITB) theory, mostly to establish the notation and symbols for the fitted parameters. Full details can be found in Refs. [5, 6, 8, and 9]. In each paper we will give a description of the particular fitting strategy used. The outline of the remaining paper is as follows. Model parameters are presented in Sec. II and compared to the previous model of Ref. 4. Sec. III describes computational setup and lists typical parameters used in simulations. In Secs. IV–VII we discuss results obtained with the new model: for water monomer and dimer (Sec. IV), hexamers (Sec. V), ice (Sec. VI), and liquid water (Sec. VII). In the latter section, in particular, we present the temperature dependence of the radial distribution functions (RDF), self diffusion coefficient, and dielectric constant of liquid water. Methanol and methanol–water mixtures are described in Sec. VIII. Sec. IX presents the main conclusions of the article.

I. FITTING STRATEGY

It was established by some of us in the course of the previous study[4] that the self-consistent polarizable ion TB provides a natural framework for the description of water able to capture all essential interactions. Even our "intuitive" models (point charge and dipole) were sufficiently good to reproduce various properties of water. However, the proper enhancement of molecular dipoles in going from water monomer to hexamers and further to ice and liquid water, could only be achieved in the polarizable ion model. Therefore, in Ref. 4 we took the

intuitive dipole model and further adjusted its parameters numerically using Schwefel's genetic algorithm.[10] In the rest of the paper, we shall be referring to this model as the genetic dipole model of Ref. 4, or simply as the TB model of Ref. 4.

Most of results obtained with the genetic dipole model of Ref. 4 were in good agreement with experiment and accurate quantum chemistry calculations. However, we were not entirely happy with three features of the model, namely: (1) the model predicts ice to be denser than liquid water; (2) there appears a shoulder near the first peak of the O–O RDF absent on experimental curves; and (3) the energy separation between oxygen s and p states is insufficient to render correct band structure of metal oxides. Hence, one cannot simulate systems containing water *and* metal oxide without having two distinct oxygen species. Interestingly enough, the point charge model of Ref. 4 did not show any of the above deficiencies, but the point charge model was inferior to dipole models in reproducing dipole moments and the monomer polarizability. That led us to think that features (1)–(3) are not unavoidable and can be eliminated in the dipole model. In the present work we achieved exactly that: developed a new dipole model for water free from undesirable features (1)–(3) and thus suitable for simulations of condensed phases of water and water—metal oxide interfaces.

The fitting procedure was organized as follows. At the first stage we concentrate on quantities that do not depend on the O–O pair potential. For these, we employed the genetic algorithm to fit on site energies, Hubbard U's, polarizability parameters Δ's, O–H and O–O hopping integrals, and O–H pair potential. The targets at this stage were mostly the monomer properties (geometry, vibrational force constants, cohesive energy, and the HOMO–LUMO gap). We also used dimer angles α and β (see Fig. 1a) at the experimental O–O distance $R_{\rm OO}^{eq} = 2.912$ Å[11] as well as derivatives of the angles with respect to the O–O distance, $d\alpha/dR_{\rm OO}$ and $d\beta/dR_{\rm OO}$, evaluated at $R_{\rm OO}^{eq}$. The derivatives of the dimer angles were not used in Ref. 4, but since in the present study we do not insist on reproducing $R_{\rm OO}^{eq}$ exactly, it becomes important to have reasonable angles away from $R_{\rm OO}^{eq}$.

Another new quantity that we added to the objective function was the energy profile corresponding to proton transfer across the dimer. The proton transfer is of direct relevance to the hydrogen diffusion in liquid water (the Grotthuss mechanism). Also, it has been discovered in quantum chemical calculations that the water dimer changes its structure at oxygen–oxygen separation shorter than 2.65 Å.[12] This feature was included into the fitting procedure as well since such re-orientation of water molecules might affect the first peak of the O–O RDF in liquid water.

The whole set of parameters found at the first stage is passed to the second stage at which we determine the O–O pair potential. For that, we no longer use the energy vs. O–O distance curve in water dimer. Instead, we fit O–O RDF in liquid water directly.

This is a lengthy procedure as every calculation of RDF requires an MD simulation 40–45 ps long. This is the reason why one wants to separate O–O pair potential from the rest of the model parameters in the first place. Genetic algorithm is not of much help here, instead we generate a few instances of the O–O pair potential, obtain the respective RDFs and then try to "interpolate" between them in order to match the experimental RDF as closely as possible. The pair potential parameters that we thus guess, are used to generate another set of RDFs, and so on. Typically, it takes 3–4 iterations to arrive at a reasonable RDF.

Usually at this stage the density of water is already within 10% of the experimental one. To improve the density further, we fix the repulsive part of the O–O pair potential and change only its attractive part. Such "fine tuning" is possible due to the fact that density is much more sensitive to the attractive part of the O–O pair potential than RDFs. As before, we create a few replicas, perform NPT for each of them to estimate respective densities, and "interpolate" into experimental density. This completes the whole cycle of the fitting.

Table I: Parameters of present TB model of water compared to those of the genetic dipole model of Ref. 4. See Ref. 1 for the meaning of parameters and abbreviations. Notation for the functional form of the scaling law for hopping integrals and pair potential is clarified in Table II. All values are given in atomic Rydberg units.

	Present	Ref. 4
On site parameters:		
ε_s(H)	–1	–1
ε_s(O)	–2.1164	–1.51
ε_p(O)	–1.1492	–1.20
U(H)	1	1.08
U(O)	1.0775	1.16
Δ_{spp}	–0.9430	–0.924
O–H hopping integrals:		
Functional form	GSP	GSP
$A_{ss\sigma}$	-0.502	–0.348
$A_{sp\sigma}$	0.436	0.313

Table I – *Continued from previous page*

	Present	Ref. 4
$n_{ss\sigma}$	2.096	1.48
$n_{sp\sigma}$	1.502	1.98
n_c	4.056	4.04
r_0	1.8094	1.8094
r_c	3.799	3.47
Cutoff $[r_c^1, r_c^2]$	[2.1, 5.5]	—
O–H pair potential:		
Functional form	GSP	GSP
A	0.73669	0.552
n	3.3502	3.362
n_c	6.3096	6.04
r_0	1.8094	1.8094
r_c	3.3550	3.04
Cutoff $[r_c^1, r_c^2]$	[2.1, 5.5]	—
O–O hopping integrals:		
Functional form	GSP	GSP
$A_{ss\sigma}$	–0.015	–0.080
$A_{sp\sigma}$	0.002	0.050
$A_{pp\sigma}$	0.050	0.00012
$A_{pp\pi}$	–0.020	–0.004
$n_{ss\sigma}$	2	2
$n_{sp\sigma}$	2	2
$n_{pp\sigma}$	3	3
$n_{pp\pi}$	3	3
n_c	6	4
r_0	5.6	5
r_c	9.0	6.8
Cutoff $[r_c^1, r_c^2]$	[8, 11]	—
O–O pair potential:		
Functional form	EPL	QUAD
A_1	4.0306×10^{-3}	0.010
m_1	10	—
p_1	0	—
A_2	-2.0265×10^{-3}	0.647
m_2	6	—
p_2	0	—
r_0	5.6	5.992
r_1	—	5.494
r_c	—	6.110
Cutoff $[r_c^1, r_c^2]$	[8, 11]	—

II. PARAMETERS OF THE MODEL

The complete set of parameters of the new water model is given in Table I. The parameters of the previous TB water model[4] are also shown in the Table for comparison. Notation for scaling laws used in Table I ("GSP", "EPL", or "QUAD") is clarified in Table II.

There are three major changes in the present model compared to the previous model. One of them, already discussed in Sec. I, is an additional r^{-6} attractive term in O–O pair potential. One can think of it as the term providing unaccounted attraction between water molecules due to dispersion and induction forces. However, here we treat it as just another parameter of the model and use it for fitting rather than pick up some predefined value (as is done in recent "DFT+D" schemes[15]). This is consistent with our empirical tight binding approach in which we search for model parameters that provide better agreement with experiment rather than reproduce DFT results.

The second difference between the present and previous models is the change of the $s-p$ splitting for oxygen. Indeed, in the present model the difference between the on site energies ε_p(O) and ε_s(O) is about 1 Ry and is close to the value that one would obtain using Hartree–Fock term values (ε_s = –2.142 Ry and ε_p = –1.038 Ry). In the previous model, the $s-p$ difference is only 0.3 Ry and leads to an incorrect valence band in metal oxides.[7]

As discussed in Ref. 4, the reduction of the $s-p$ splitting leads to the increase of the bond angle in a wa-

Table II. Explicit form of the scaling laws referred to in Table I.

Notation	Function	Explicit form
GSP	Goodwin–Skinner–Pettifor[13]	$f(r) = A\,(r_0/r)^n \exp\{n\,[-(r/r_c)^{n_c} + (r_0/r_c)^{n_c}]\}$
EPL	Exponential × power law	$f(r) = \sum_i A_i\,(r_0/r)^{m_i}\,\exp\,[-p_i(r-r_0)]$
QUAD	Chadi's quadratic[14]	$f(r) = A_1\epsilon + A_2\epsilon^2$, where $\epsilon = (r-r_0)/r_0$

ter molecule from 90° (which it would have had in the limit of the large splitting) towards its target value of 104.51° (see Table III). By imposing a lower bound on the $s-p$ splitting, we forced the genetic algorithm to search for other means to achieve the required bond angle, namely via adjusting the O–H hopping integrals and on site Hubbard–U parameters. As we shall see, the resulting model appears to outperform our previous model, which might seem paradoxical given that the new model is derived under additional constraints. The reason for that, to our belief, is the fact that the constraint drove the system into a proper, more physical basin of the phase space of parameter values.

The third modification of the model is the introduction of explicit cutoffs into all distance dependences.[16] Given that the present model is intended for simulation of condensed phases, it is essential that at specified distances both pair potentials and hopping integrals smoothly turn to zero. Otherwise, long MD runs might be blighted with energy leakage and even become unstable.

In addition to the above three changes, we introduced a short range repulsive H–H pair potential. There are no other interactions between hydrogen atoms in our model since these are considered insignificant. The weak repulsion is added with the sole purpose of avoiding polarization catastrophes. Two H atoms that happen to be close enough during a simulation, may spontaneously pick up opposite charges. That would lead to the pair being pulled toward each other by resulting electrostatic attraction. A repulsive pair potential diverging faster than $1/r$ prevents this from happening (here, we used $1/r^{12}$).

III. CALCULATION DETAILS

TB calculations are performed with our empirical tight binding (TBE) computer code in which the self consistent polarizable ion method is implemented. All the results reported below are obtained using the orthogonal TB method[19] without spin polarization.

For MD simulations we employ reversible integrators with Liouville operators.[20] A single Nosé–Hoover thermostat is used for the temperature control, to which the particles and the barostat are coupled with relaxation times of 0.1 ps and 0.4 ps, respectively. To ensure good energy conservation, a small time step of 0.5 fs is chosen

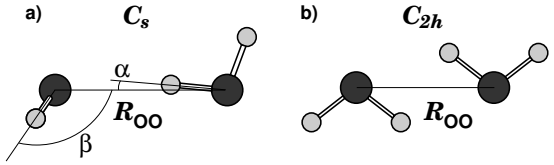

Figure 1. Two structures of water dimer: a) H–bonded dimer, point group C_s; and b) planar dimer with antiparallel dipoles, point group C_{2h}. The latter has lower energy at short R_{OO},[12] whereas the former provides the global energy minimum.

in canonical (NVT) and microcanonical (NVE) simulations, whereas in isothermal–isobaric (NPT) runs it is further reduced down to 0.25 fs. Note that we do not replace hydrogen with deuterium, all the results refer to the light rather than heavy water. A typical MD simulation consists of 20 ps of equilibration followed by >100 ps of production run.

NVT simulations of liquid water at different temperatures in Sec. VII are performed with 128 water molecules in a cubic box of the size corresponding to a density of 1 g/cm^3, and an orthorhombic cell containing 8 water units (24 atoms) is used for ice (Sec. VI).

Computations of water monomer, dimer, and hexamers are non-periodic, whereas in periodic condensed phase simulations we use the 12×6×6 $k-$point mesh (ice) or only the Γ-point (liquid water) to sample the Brillouin zone.

Additional quantum chemistry calculations are performed at the MP2 and CCSD(T) levels as implemented in the GAMESS computer package.[21] An augmented, correlation-consistent, valence polarization triple zeta (aug-cc-vPTZ) basis set is used throughout with the basis set superposition error corrected.

IV. WATER MONOMER AND DIMER

Tables III and IV present some properties of the water monomer and dimer as given by our new TB model and by the previous model of Ref. 4. All of these are the fitting results (with the exception of atomic charges δ in Table III), for which the last line in the Tables lists the target values.

As discussed in Sec. I, we do not put as much emphasis on monomer and dimer as we did in Ref. 4. Therefore it

Table III. Properties of an isolated water molecule: partial charge on H atom (δ), equilibrium O–H distance ($R_{\rm OH}$) and HOH angle (θ), vibrational force constants: symmetric stretch (ν_1), bending (ν_2), and asymmetric stretch (ν_3), polarizability ($\alpha_{\rm H_2O}$), dipole ($\mu_{\rm H_2O}$), HOMO–LUMO gap ($E_{\rm gap}$), and cohesive energy ($E_{\rm coh}$). Target values are experimental data from the CRC Handbook[17] unless indicated otherwise.

	δ	$R_{\rm OH}$	θ	ν_1	ν_2	ν_3	$\alpha_{\rm H_2O}$	$\mu_{\rm H_2O}$	$E_{\rm gap}$	$E_{\rm coh}$
	($\lvert\bar{e}\rvert$)	(Å)	(deg.)	(Ry/Bohr^2)			(Å^3)	(D)	(Ry)	(Ry)
TB (present)	0.46	0.9580	104.46	1.037	0.092	0.935	1.348	1.843	0.813	0.858
TB (Ref. 4)	0.47	0.9575	104.26	1.029	0.065	1.061	1.470	1.858	0.660	0.755
Target		0.9575	104.51	1.029[a]	0.100[a]	1.062[a]	1.45	1.855±0.005		0.911[b]

[a] Ref. 18.

[b] CCSD(T) results, present study. Target value for cohesive energy $E_{\rm coh} = 0.911$ Ry is the CCSD(T) cohesive energy of 0.735 Ry less the spin-polarisation energy of O atom $E_{\rm sp-pol} = -0.176$ Ry (see text).

Table IV. Properties of the water dimer: equilibrium O–O distance $R_{\rm OO}$, dimer angles α and β (see Fig. 1a), and dissociation energy $E_{\rm diss}$. Target values are CCSD(T) results by Klopper *et al.*[11]

	$R_{\rm OO}$ (Å)	α (deg.)	β (deg.)	$E_{\rm diss}$ (mRy)
TB (present)	2.7851	2.3	124.9	–15.8
TB (Ref. 4)	2.9153	3.0	113.7	–15.1
Target	2.912±0.005	5.5	124.4	–15.9

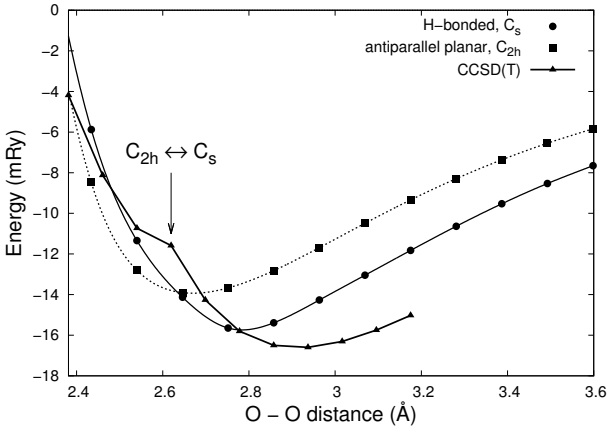

Figure 2. Binding energy of water dimer: H–bonded, point group C_s (circles) and planar with antiparallel dipoles, point group C_{2h} (squares). The CCSD(T) curve is shown with triangles, the kink on the curve indicates the transition between C_s and C_{2h} (shown with an arrow). According to Burnham and Xantheas,[12] a non-planar version of C_{2h} with point group C_i becomes lowest in energy between 2.5 Å and 2.66 Å. However, we find unnecessary going into these fine details for fitting purposes and required only that the $C_{2h} \leftrightarrow C_s$ transition is reproduced.

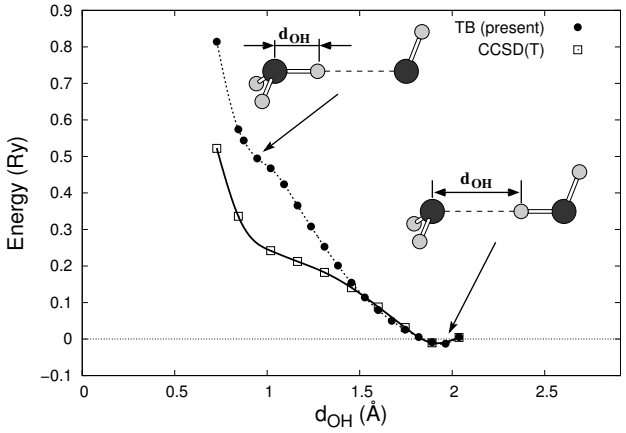

Figure 3. Binding energy profile corresponding to the transfer of a proton in water dimer as given by the present TB model (circles) and by CCSD(T) (squares). The dimer is in configuration C_s (Fig. 1a) at experimental $R_{\rm OO}^{eq} = 2.912$ Å. The oxygen atom of the acceptor molecule is situated at the origin, oxygen atom of the donor molecule is at the right end of the plot (2.912 Å). The proton moves along the O–O line from the donor molecule toward the acceptor molecule. Positions of the other atoms are fixed. Atomic configurations near the both ends of the curve are shown as two insets.

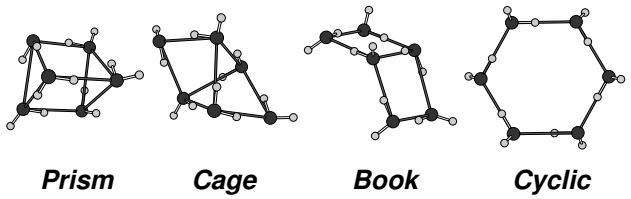

Figure 4. Atomic configuration of water hexamers discussed in Sec. V: the prism, the cage, the book, and the planar cyclic hexamer (the ring). The number of hydrogen bonds in these hexamers is 9, 8, 7, and 6, respectively.[22]

is not surprising that some of results worsen compared to the previous model. In particular, the present model fails to reproduce the peculiar ordering of force constants in the monomer $\nu_1 < \nu_3$. In addition, the present model strongly underestimates the equilibrium O–O distance in the water dimer. Other results represent some improvement over the previous model, such as the bending force constant ν_2 and the HOMO–LUMO gap for the monomer, and the dissociation energy of the dimer $E_{\rm diss}$.

The cohesive energy $E_{\rm coh}$ of the monomer deserves a special comment. The CCSD(T) result 0.735 Ry is close to experimental atomization energy at 0 K, $E_{\rm at} =$ 0.6992 Ry,[23] so either of these values could be used as a target. However, they both assume that H_2O splits into isolated atoms *in their ground state*, which for O in particular implies the spin-polarized state with multiplicity 3. Since in our non-spin-polarized calculations oxygen can only be paramagnetic, we correct the above $E_{\rm coh} = 0.735$ Ry by the spin-polarization energy of the oxygen atom $E_{\rm sp-pol} = -0.176$ Ry and arrive at 0.911 Ry as a more suitable target value for $E_{\rm coh}$.

Figs. 2 and 3 refer to additional features included into fitting: re-orientation of the dimer at short O–O distance (Fig. 2) and the energy profile arising while the H–bond hydrogen is being pulled across the dimer (Fig. 3). The TB curves in both cases are compared against the CCSD(T) curves computed in the present study.

A change of the relative orientation of molecules in a water dimer at O–O separation less than 2.66 Å is predicted on the basis of MP2 calculations by Burnham and Xantheas.[12] In our own computations (MP2 relaxation followed by the CCSD(T) total energy calculation) we also found that configuration C_{2h}, Fig. 1b, becomes lower in energy than configuration C_s, Fig. 1a, at 2.62 Å. This energy crossover was added to the list of properties to be fitted (see Sec. I) since it might affect water–water interaction at short distances and therefore the shape of the first peak in O–O RDFs. As one can appreciate from Fig. 2, the re-orientation distance is reproduced in the present TB model rather accurately.

The proton transfer curve in Fig. 3 was included into fitting as this is a way to sample the hydrogen bond profile. The curves in the Figure were obtained by placing the hydrogen atom on the O–O line at various distances $d_{\rm OH}$, with all other atoms frozen in their equilibrium positions in the C_s dimer, Fig. 1a. The obtained TB curve closely follows the CCSD(T) curve in the vicinity of the hydrogen equilibrium and up to the half way towards the acceptor molecule. From there, the two curves begin to deviate. However, this is not too alarming since the curves appreciably differ only at energies hardly accessible in an ordinary MD run.

The proton transfer curve can also be linked to the water self ionization reaction

$$2H_2O \rightleftharpoons H_3O^+ + OH^- + Q\,, \qquad (1)$$

except that the reaction products are not separated to infinity. That is the reason for the absence of the second, local minimum near the high energy end of the curve. The thermal effect of reaction (1), however, can easily be obtained by combining energies of the isolated molecules. The TB result is $Q = -0.929$ Ry (-610 kJ/mol H_2O) compared to CCSD(T) result of $Q = -0.719$ Ry (-472 kJ/mol H_2O). It refers to water self ionization in vacuum at 0 K without any zero point energy correction. Although there is a 30% difference between the TB and CCSD(T) results, what is important is that charged molecules are satisfactorily dealt with in the present model, and that the large energy of the reaction (1) is reproduced.

It is also essential that neither of the above curves demonstrates any pathological features at short O–O separation. In the previous model,[4] on the contrary, there was a sudden energy jump at the proton transfer curve as the proton approached the acceptor molecule. It turned out that the three H and one O atoms formed an "anti-hydronium" molecule, with oxygen charged positively and hydrogens charged negatively. Pushing the O$-p$ state down relative to the H$-s$ state was sufficient to cure the problem.

Table V. Structure and energies of the four lowest energy water hexamers: prism, cage, book, and the planar cyclic hexamer depicted in Fig. 4. $R_{\rm OO}$ is the distance between adjacent oxygens (in the planar cyclic hexamer, the length of all six O–O bonds are the same), $E_{\rm diss}$ is the energy required to dissociate a hexamer into six water molecules, and $\Delta E_{\rm diss} = E_{\rm diss}$(hexamer) – $E_{\rm diss}$(prism) are the relative energies of each hexamer with respect to the prism hexamer. Target energies are the CCSD(T) results obtained using the aug-cc-pVTZ basis set by Olson *et al.*[24] The target $R_{\rm OO}$ distance in the planar cyclic hexamer is an MP2 result by Santra *et al.*[25]

Property:	$R_{\rm OO}$ (Å)	$E_{\rm diss}$ (Ry)	$\Delta E_{\rm diss}$ (mRy)			
Hexamer:	cyclic	prism	prism	cage	book	cyclic
TB (present)	2.7346	–0.141	0.0	1.7	7.0	13.0
TB (Ref. 4)	2.6506	–0.144	0.0	2.8	6.9	11.8
Target	2.7069	–0.154	0.0	0.88	3.87	6.82

V. WATER HEXAMERS

Water hexamers are the natural next objects to test the performance of our model aimed for the condensed phase simulations. Firstly, this is because the water–water bonding in liquid water and ice is much closer to that in hexamers than in the dimer. Secondly, water hexamers are perhaps the biggest water clusters that have been systematically studied in the literature, including accurate *ab initio* computations.[22,24,25]

In Table V we compare results obtained with our present and previous water models to *ab initio*. The

agreement between all three sets of data is very good. Importantly, both TB models give the correct energy ordering of the isomers: prism < cage < book < cyclic. The ordering is commonly believed to arise from the variation in the number of hydrogen bonds: 9 for the prism, 8 for the cage, 7 for the book, and 6 for the cyclic hexamer,[22] although, as noted by Santra *et al.*,[25] there is a certain ambiguity in this argument.

Overall, we are satisfied with the predictions of our TB model for hexamers. Note that hexamers are not included into the fitting set, and neither is ice which we consider next.

VI. ICE XI

The most common form of ice that we see outdoors in winter or inside an old fridge in the kitchen is the hexagonal ice, Ih. Hexagonal ice is a proton disordered phase, but at low temperature (72 K for H_2O, 76 K for D_2O) the protons order, and ice Ih converts to the ferroelectric orthorhombic phase ice XI which is believed to be the lowest energy modification of ice at ambient pressure and 0 K. According to neutron diffraction experiment the space group of ice XI is $Cmc2_1$.[30,32]

In a recent study Hirsch and Ojamäe[26] considered 16 proton ordered structures that are possible to arrange in a 24 atom orthorhombic unit cell, including ice XI. It was found that density functional calculations with different types of basis sets and different flavors of generalized gradient approximation (GGA) do favor the ice XI structure, whereas force field models predicted a different phase with $P2_12_12_1$ space group symmetry ("phase number 6" in Ref. 26) to be the lowest in energy.

Using our TB model we optimized the unit cell parameters and atomic coordinates of the both ice polymorphs, ice XI and the hypothetical phase of Ref. 26. We also found the former to be lower in energy, as with the previous water model of Ref. 4. In Table VI we list the lattice energies and densities of ice obtained with TB along with those found in gradient corrected DFT calculations using GGA (PW91, PBE, BLYP), hybrid functionals (B3LYP, PBE0), meta–GGA (M06–L), and dispersion corrected functionals (BLYP+D). Judging from the Table, there is no clear advantage in a particular type of exchange-correlation functional. Perhaps, B3LYP and M06–L provide a somewhat better agreement with experiment. The tight binding results appear in this respect at least as good as the DFT ones. Besides, in the present TB model the liquid water is denser than the ice, which was an essential requirement to the model (see Sec. I).

VII. LIQUID WATER

We move now to the model predictions for the liquid water. This is the most important part of the results as liquid water is where the majority of the model application is expected to lie.

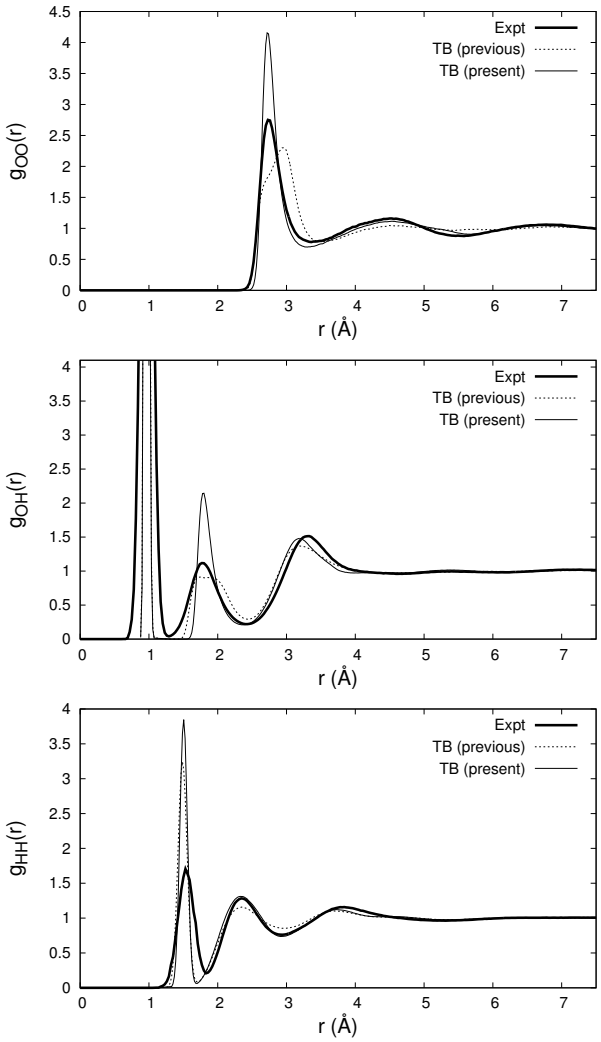

Figure 5. Radial distribution functions $g(r)$ in liquid water for O–O, O–H, and H–H neighbors obtained with the present TB model (thin line) and with the "genetic" dipole model from Ref. 4 (dotted line). The experimental curves (thick line) are from Refs. 45 and 46. Theoretical curves correspond to $T = 300$ K and zero pressure, whereas experimental curves are taken at $T = 298$ K and $P = 0.1$ MPa.

The density of water at ambient conditions was evaluated after an extended NPT run with 50 ps equilibration time and 118 ps of the production run at zero pressure and $T = 300$ K. The resulting average density is $\bar{\rho} = 1.008 \pm 0.026$ g/cm^3 at average temperature $\bar{T} = 300 \pm 12$ K (the error bars are estimated from the dispersion of a quantity in question). Hence, we use $\rho = 1.0$ g/cm^3 in all subsequent NVE and NVT room temperature simulations.

Table VI. Lattice energy $E_{\rm lat}$ and density ρ of ice XI at 0 K. "Hypothetical" phase is "phase number 6" from Ref. 26 which was found to be favored over ice XI in force field models. $\Delta E_{\rm lat}$ is the difference between lattice energies of the hypothetical phase and ice XI. Densities of liquid water are also shown for comparison.

	Ice XI (space group $Cmc2_1$)		Hypothetical (space group $P2_12_12_1$)		Liquid
	ρ (g/cm^3)	$E_{\rm lat}$(kJ/mol)	ρ (g/cm^3)	$\Delta E_{\rm lat}$(kJ/mol)	ρ (g/cm^3)
TB (present)	0.968	–54.26	0.940	0.21	1.008
TB (Ref. 4)	0.967	–51.08		0.67	0.926
DFT–GGA (PW91)	0.995[a]	–68.66[a]	0.969[a]	0.96[a]	
DFT–GGA (BLYP)		–55.85[b]		0.71[b]	0.92[c]
DFT–GGA (BLYP+D)		–69.434[e]		0.616[e]	1.07[c]
DFT–GGA (PBE)	0.989[d]	–66.74,[d] –67.901[e]		0.571[e]	
DFT–GGA (M06–L)	0.953[d]	–61.04[d]			
DFT–GGA (B3LYP)	0.940[d]	–59.29[d]			
DFT–GGA (PBE0)	0.981[d]	–63.76[d]			
Experiment	0.935[f]	–58.87,[g] –59.25[h]		> 0	0.997

[a] Ref. [26].
[b] Ref. [26], structural optimisation only.
[c] Ref. [27].
[d] Ref. [28].
[e] Ref. [29]. Results slightly depend on the computer program used. We cite only those obtained with the CP2K code.
[f] Measurements at 5 K.[30]
[g] Experimental estimation at 0 K with zero point energy removed, see Refs. 28 and 31 and references therein.
[h] Estimated using the lattice energy of ice Ih and the enthalpy of the ice Ih $\longrightarrow$ ice XI transition, see Ref. 29.

Table VII. Properties of liquid water predicted in our present and previous (Ref. 4) TB models at ambient conditions (zero pressure, $T = 300$ K): N_c is the average O–O coordination number, $\bar{p}_{\rm tot}$ and $\bar{p}_{\rm ind}$ are the average total and induced dipole moments of a water molecule, ε_0 is the static dielectric constant, and $D_{\rm self}$ is the self diffusion coefficient. TB values for the present model are interpolated into $T = 300$ K using the whole set of data between 260 K to 350 K. Available experimental and DFT–GGA data are also shown for comparison. Diffusion coefficients in parenthesis correspond to D_2O rather than H_2O.

	N_c	$\bar{p}_{\rm tot}$ (D)	$\bar{p}_{\rm ind}$ (D)	ε_0	$D_{\rm self}$ (10^{-5} cm^2/s)
Experiment	4.67±0.05[a]	2.95±0.2[b]	–	78[c]	2.23±0.1[d] (1.9[e])
TB (present)	4.54	3.15	0.67	80.0	1.98
TB (Ref. 4)		3.28	0.74	86.7	3.0
DFT–GGA (PBE)		2.95,[f] 3.09[g]	–	67±6,[g] 75[h]	(1.6[g])
DFT–GGA (BLYP)					0.25,[i] 0.55[j]
DFT–GGA (BLYP+D)					(1.7[k])
DFT–GGA (DRSLL)	4.90[l]				(2.63[l])
DFT–GGA (DCACP)					(2.1±0.23[m])

[a] 298 K, Ref. [33].
[b] 300 K, Ref. [34].
[c] 300 K, Ref. [35].
[d] 298 K, Ref. [36].
[e] 298 K, Ref. [37].
[f] 318 K, Ref. [38].
[g] 330 K, Ref. [39].
[h] Extrapolated to 300 K from data in Ref. [39].
[i] 308 K, Ref. [40].
[j] 305 K, Ref. [41].
[k] 317 K, Ref. [42].
[l] 300 K, Ref. [43].
[m] 325 K, Ref. [44].

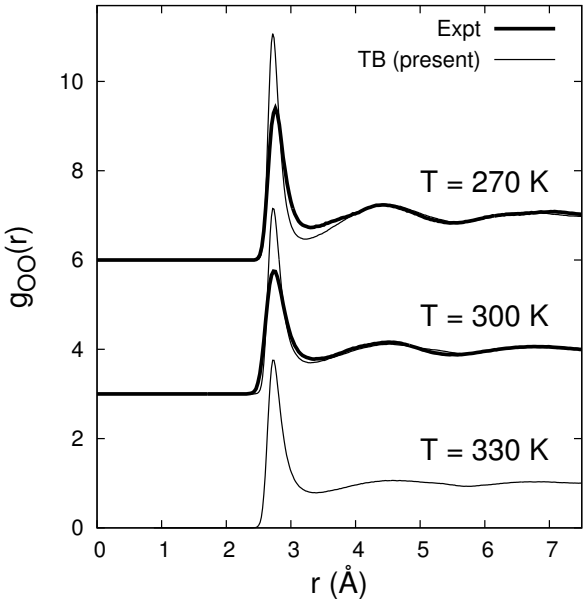

Figure 6. O–O radial distribution function at different temperatures. TB curves (thin lines) are obtained at $T = 270$, 300, and 330 K and zero pressure. Experimental curves (thick lines) are measured at $T = 298$ K, $P = 0.1$ MPa and $T = 268$ K, $P = 27$ MPa.[45,46] Subsequent sets of curves are shifted vertically by 3 units for clarity. RDFs at ambient conditions (the middle set) repeat curves in the top panel in Fig. 5.

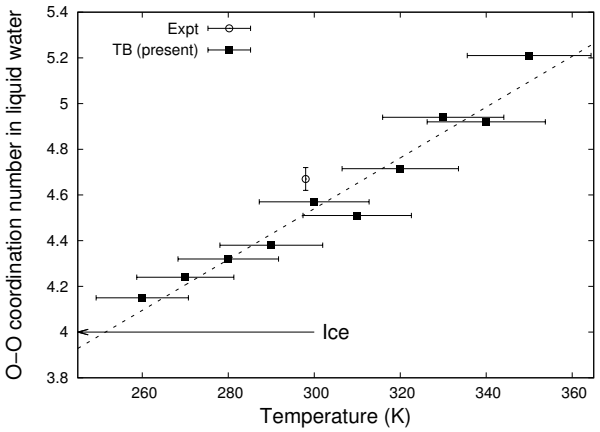

Figure 7. Temperature dependence of the O–O coordination number N_c in liquid water: TB results (squares) and experiment (circle).[33] Dotted line is the linear fit of $N_c(T)$ and is used to obtain the coordination number at room temperature (see Table VII). Temperature error bars shown for the TB set are estimated from the dispersion of instant temperature. The horizontal arrow indicates the coordination number in ice, which is always 4 apart from amorphous and some exotic high pressure phases.[47]

A. Radial distribution functions and the "fifth neighbor"

In Fig. 5 we show the radial distribution functions (RDFs) $g(r)$ obtained after NVT simulation at $T = 300$ K and $\rho = 1.0$ g/cm^3 together with those obtained with the previous model[4] and by neutron diffraction.[45]

The theoretical O–O curve (thin solid line in the top plot of Fig. 5) provides an excellent match for the position of all three maxima on $g_{OO}(r)$ and reasonable match for the position of two minima between them. The first peak is more localized than in experiment but is very well shaped, which we consider as a significant improvement over the one predicted by our previous model with a shoulder and displaced maximum (dotted line in the same figure). A nice agreement of the O–O RDF is not just a coincidence, of course, as it was included into the fitting procedure (see Sec. I).

The O–H and H–H RDFs were not directly fitted, hence it is not surprising that the agreement with experiment for them is less perfect. The O–H RDF in particular seems farther from experiment than RDF of the previous model, whereas the H–H RDF appears again a better match. In any case, it is safe to conclude that the present model predicts correct peak positions and provides well defined shapes for these peaks on all three types of radial distribution curves.

Evolution of the oxygen–oxygen RDF with temperature is shown in Fig. 6 together with available experimental RDFs. We observe that $g_{OO}(r)$ at $T = 270$ K, again, has a more confined first peak than its experimental counterpart, but starting from the second peak the TB curve remains in the perfect agreement with experiment. Increasing the temperature smears the structure of $g_{OO}(r)$, although there appears to be a bigger change in going from 270 K to 300 K, than from 300 K to 330 K. The smearing means that the maxima decrease in height and widen; as a result the first minimum moves to the right. In addition, the second maximum slightly moves to the right too, as is also noted in other simulations (see Ref. 48 and references therein).

One of the characteristic features of the structure of liquid water is a non-integer coordination number N_c of a water molecule. If water had had a perfect tetrahedral coordination then N_c would have been equal to four (as is the case of ice XI). Experiment however predicts a fractional number between 4 and 5. For instance, a recent estimation by Soper and Benmore[33] gives $N_c = 4.67 \pm 0.05$. The reason behind N_c being larger than four is a subject of extensive discussion in the literature, and is often referred to as the "fifth neighbor" problem.[42]

In our TB model we obtain $N_c = 4.54$, in good agreement with the experimental estimation by Soper and Benmore (see Fig. 7 and Table VII). One might argue that this agreement might simply be the consequence of $g_{OO}(r)$ being part of the fitting. We do not think so since our first peak differs in shape from the experiment to such an extent that the chances for it to have exactly the same area are slim.

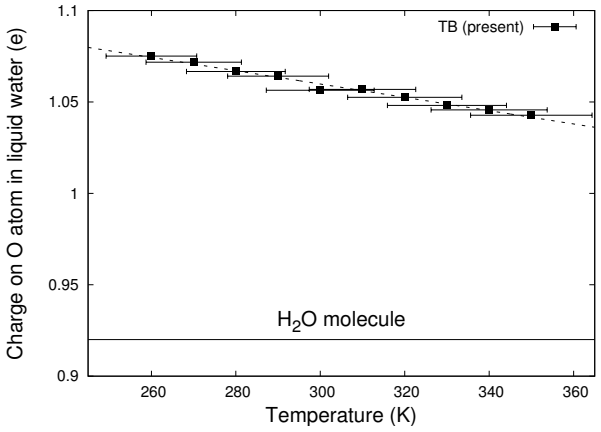

Figure 8. Average charge on oxygen atom $q_{\rm O}$ (in electrons) in liquid water as a function of temperature. Dotted line gives the linear fit of $q_{\rm O}(T)$ dependence. Horizontal line corresponds to oxygen charge in isolated water molecule (Table III).

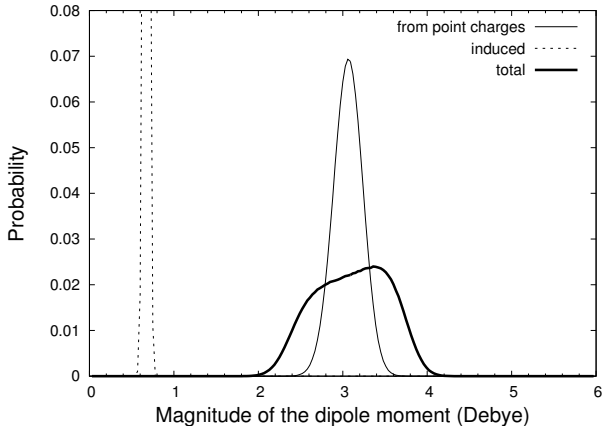

Figure 9. Distribution of the magnitude of the dipole moment of water molecules at $T = 300$ K. Dipole due to point charges (thin line) and induced dipole of oxygen atoms (dotted line) combine vectorially together to form the total dipole moment magnitude (thick line). The point charge dipole is increased compared to that of water monomer ($\mu_{\rm pc} = 2.60$ D), and the average total dipole is increased even further (see Sec. VII B for discussion).

However, we might have a good N_c for a different reason, namely, because the *density* of liquid water is included into the fitting. As a matter of fact, we kept tuning the strength of our "dispersion term" until the density of liquid water became 1 g/cm^3 (see Sec. I), whereas according to Wang *et al.*,[43] it is the dispersion interaction that appears responsible for the coordination number.

Given the weakness of the dispersion forces, it is not surprising that the temperature dependence of N_c is so strong that N_c already approaches that of ice at the left edge of the plot in Fig. 7. The $N_c(T)$ dependence of supercooled liquid water is perhaps weaker than the linear fit for the whole temperature interval (shown in the Figure with the dotted line). The interesting observation, however, is the fact that the coordination numbers of liquid water and ice approach each other as the temperature decreases towards the temperature of homogeneous nucleation of water ($T_H \simeq -38^\circ$C at 1 bar) below which water cannot stay liquid.[49]

B. Why is the water dipole moment bigger in liquid than in a monomer?

One of the important properties of liquid water is the fact that the dipole moment of a water molecule increases from 1.86 D (see Table III) to 2.95 D (see Table VII) as molecules gather to form a liquid. What is the reason, or perhaps the physical mechanism, behind the enhancement of the dipole?

Although dipoles in condensed matter are known to be ill-defined, it is still a useful quantity in many respects. For example, a big molecular dipole in liquid water is often related to its large dielectric constant, which is observable.

The tight binding formalism represents a perfect framework for investigations of this type. One does not even need to introduce Wannier functions or special projector operators to define point charges or dipoles (as in DFT), they are already a part of the TB formalism.

Returning to the question posed in the title of this Section, it appears one can distinguish three sources of dipole enhancement in water.

First of all, as water molecules combine together and form hydrogen bonds, there must be some additional redistribution of charge between the species. In TB terms, this is to say that the point charges of species change. Indeed, Fig. 8 shows a substantial increase of the average charge on oxygen q_O (average charge on hydrogen is $q_H = -q_O/2$ as the whole system is neutral). Given that the geometry of water molecules does not change much, the average dipole that these point charges create, $\bar{p}_{\rm pc}$, increases as well (we refer to this dipole as the *point charge* dipole, to distinguish it from an induced dipole $\bar{p}_{\rm ind}$ that arises due to atomic polarizability). In our calculations, the point charge dipole in liquid water at room temperature is $\bar{p}_{\rm pc} = 3.08$ D, and is larger than in water monomer $\mu_{\rm pc} = 2.60$ D.

The second source of the dipole enhancement comes from the incomplete compensation of the point charge and induced dipole moments. Indeed, in a single water molecule the point charge dipole $\mu_{\rm pc} = 2.60$ D mentioned above polarizes the oxygen atom and induces an opposite dipole $\mu_{\rm ind} = 0.76$ D. These two dipoles combine together to make the total dipole $\mu_{\rm tot} = 1.84$ D (see Table III).

In liquid water superposed electrostatic fields destroy this maximal cancellation of the point charge and induced dipoles, hence the total dipole moment is not decreased

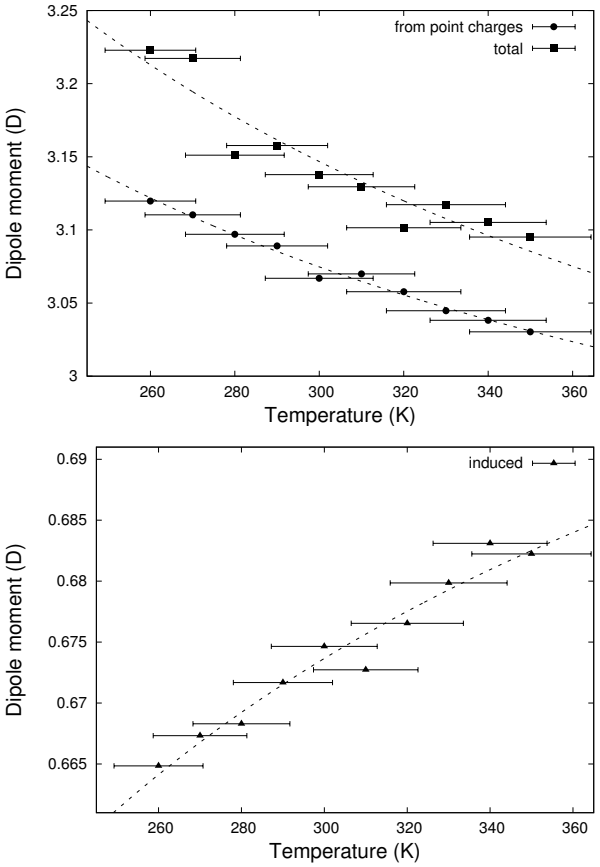

Figure 10. Temperature dependence of the total dipole moment (squares, upper panel), dipole from point charges (circles, upper panel), and induced dipole (triangles, lower panel) in liquid water. These are average moments per water molecule given by the center of gravity of respective moment distributions, such as those shown in Fig. 9. The dotted lines are the inverse temperature fits in the form $f(T) = a + b/T$ and are used to interpolate dipole moments into room temperature (Table VII).

to the extent it is in the monomer. For a more detailed description of the incomplete compensation mechanism the reader is referred to Ref. 4 (see Fig. 8 there and the discussion around it). Note that this mechanism is only at work in a dipole polarizable ion TB model, in a point charge model it does not exist. That is the reason why some non-polarizable force field models have to define a larger dipole in the monomer. For the same reason, in our previous point charge model[4] we had to assign large polarizability to a water molecule.

Fig. 9 shows perfectly symmetric distributions for point charge and induced dipole moments. The center of gravity of these distributions corresponds to the average moments $\bar{p}_{\rm pc} = 3.08$ D and $\bar{p}_{\rm ind} = 0.67$ D (see Table VII). The sum of these distributions produce an asymmetric distribution for the total dipole moment $\mathbf{p}_{\rm tot} = \mathbf{p}_{\rm pc} + \mathbf{p}_{\rm ind}$ in such a way that its average value increases: $\bar{p}_{\rm tot} = 3.15$ D (Table VII).

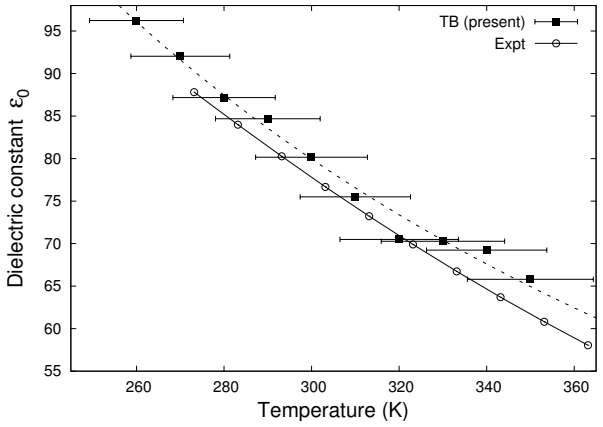

Figure 11. Temperature dependence of the static dielectric constant ε_0 given by TB (squares) and experiment (circles).[50] The dotted line and the temperature error bars have the same meaning as in Fig. 10

Thus Fig. 9 suggests that there should be a third source of dipole enhancement. Indeed, according to the first source $\bar{p}_{\rm tot}$ increases because $\bar{p}_{\rm pc}$ increases. According to the second source, $\bar{p}_{\rm tot}$ increases because $\bar{p}_{\rm pc}$ and $\bar{p}_{\rm ind}$ do not always have exactly opposite direction (as they do in monomer). This still implies that $\bar{p}_{\rm tot}$ should be smaller than $\bar{p}_{\rm pc}$ although, perhaps, not as much as in monomer. In fact, we find in Fig. 9 that $\bar{p}_{\rm tot}$ is larger than $\bar{p}_{\rm pc}$.

The "third source of dipole enhancement" is even more apparent in Fig. 10 where the total dipole and its components are shown at different temperatures. The enhancement appears nearly independent of temperature: the difference between the total and point charge dipoles stays close to 0.1 D in the whole temperature range.

This effect might seem puzzling at the first glance, but in fact it is simply a property of statistical averages. It can be shown that if $\mathbf{p}_{\rm pc}$ and $\mathbf{p}_{\rm ind}$ are uncorrelated and $\mathbf{p}_{\rm tot} = \mathbf{p}_{\rm pc} + \mathbf{p}_{\rm ind}$, then indeed $\bar{p}_{\rm tot} > \bar{p}_{\rm pc}$. Of course, $\mathbf{p}_{\rm pc}$ and $\mathbf{p}_{\rm ind}$ cannot be independent, as the latter includes polarization response to the former. Thus, we conclude that the anti-correlation between $\mathbf{p}_{\rm pc}$ and $\mathbf{p}_{\rm ind}$ is insufficiently strong to reverse the dipole enhancement. A possible reason for the decoupling of $\mathbf{p}_{\rm pc}$ and $\mathbf{p}_{\rm ind}$ could be a reduction of the point charge dipole by the electrostatic field of neighboring molecules. Another possibility is that the fifth neighbor diverts $\mathbf{p}_{\rm ind}$ away from the $-\mathbf{p}_{\rm pc}$ direction when entering the first coordination shell.

C. Dielectric constant

The large static dielectric constant ε_0 is another peculiar feature of liquid water. We calculate it in our study as

$$\varepsilon_0 = 1 + \frac{4\pi}{3}\,\frac{\langle \mathbf{P}^2 \rangle}{V k_B T},$$

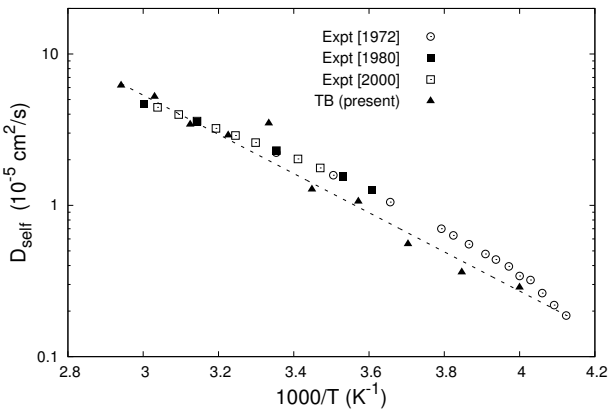

Figure 12. Arrhenius plot of the self diffusion coefficient $D_{\rm self}$ in liquid water between 260 K and 350 K obtained in TB simulations (triangles). Experimental points are from Ref. 36 (open circles), Ref. 51 (filled squares), and Ref. 52 (open squares). Dotted line represents the Arrhenius fit to TB diffusion data and is used to interpolate $D_{\rm self}$ into $T = 300$ K (see Table VII). The slope of the line corresponds to diffusion activation energy $E_a = 25$ kJ/mol.

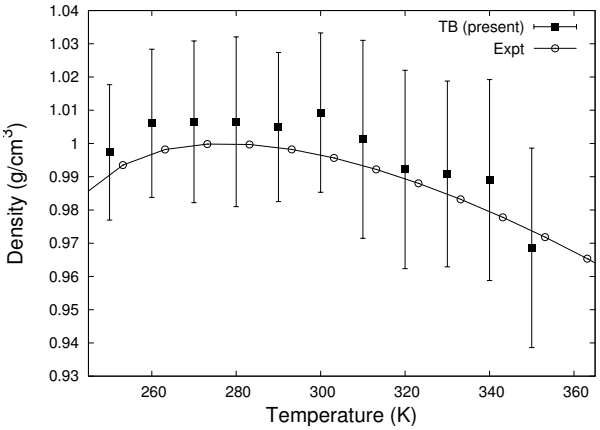

Figure 13. Temperature dependence of the density of liquid water given by TB (squares) and experiment (circles).[17] For the sake of clarity, only the density error bars are shown.

where $\mathbf{P}$ is the dipole moment of the simulation cell, V is its volume, and k_B is the Boltzmann constant. For that we employed the method proposed by Sharma, Resta, and Car[39] which relies on the radial integration of the dipole–dipole pair correlation function.

The resulting dielectric constant as a function of temperature is shown in Fig. 11 together with its least square fit in the form $\varepsilon_0(T) \simeq a+b/T$ as a dotted line. Despite a small scatter of points, the agreement between TB results (squares) and experiment (circles in Fig. 11) is remarkable in the whole temperature range.

D. Self diffusion coefficient

The same set of NPT simulations that we used to extract dielectric constant was employed again to estimate the self diffusion coefficient $D_{\rm self}$ of water. A remark on our choice of the MD ensemble seems required at this point. Being a dynamic rather than static property, $D_{\rm self}$ must depend on a particular ensemble used in MD simulation. Most diffusion simulations in literature are performed in NVE ensemble since the alternatives, such as NVT or NPT ensembles, might involve the rescaling of time variable. However, as Tuckerman pointed out,[53] it should be possible to avoid any time rescaling in both NVT and NPT simulations, which is the way the MD routine is programmed in our TBE code. This opens the possibility to simulate diffusion at specified temperature and pressure, and therefore much better represents experimental conditions. We also verified our NVT self diffusion coefficients against those obtained in a more standard NVE ensemble at temperature between 280 K and 310 K.

Our calculated coefficients of self diffusion are listed in Table VII together with experimental data and DFT results. The agreement between TB and experiment is good, much better than the agreement between experiment and DFT–GGA. It is well recognized that in order to bring DFT–GGA in agreement with experiment one needs to rescale the temperature by about 20%.[41] The situation dramatically improves, however, in the dispersion–corrected GGAs (see the BLYP+D, DRSLL, and DCACP results in Table VII).

The temperature dependence of $D_{\rm self}$ is shown in Fig. 12 together with experimental data. We use an Arrhenius plot $\log D_{\rm self}$ vs. $1/T$ in which a straight line corresponds to a thermally activated process with neither energy barrier nor attempt frequency being temperature dependent. Experiment suggests that water should be slightly non-Arrhenius: indeed, the curve is steeper at low temperatures.[36] Mills[37] provides the following estimates for the diffusion activation energy: $E_a = 20$ kJ/mol for temperatures between 1°C and 15°C, and $E_a = 18$ kJ/mol in the interval from 15°C to 45°C. It is encouraging that the TB data points represent experiment rather closely in the whole temperature range (see Fig. 12). The scatter of TB data due to the size of the system and the length of MD run is too large to resolve the curvature on the Arrhenius plot. However, if we ignore the curvature and draw a single line through data points (dotted line in Fig. 12), then we arrive at the the activation energy $E_a = 25$ kJ/mol, in excellent agreement with Mills' estimations.

E. Is density maximum of water reproduced?

One of the most important and intriguing properties of liquid water is the fact that water density has a max-

imum at 3.98°C. Important, because without this maximum life on Earth would have hardly emerged, at least in its current form. Intriguing, since the reasons behind this phenomenon are still being debated. One of the suggestions,[54] for instance, is that the maximum is a consequence of the thermodynamic singularity, "the second critical point" of water at temperature −45°C which is experimentally unreachable.

Encouraged with the excellent performance of the TB model of water seen in previous sections, especially in Secs. VII C and VII D, we decided to compare densities obtained at different temperatures to check if there is any indication of maximum. The result is shown in Fig. 13 and is inconclusive, mostly because of the large density error bars (obtained, as usual, as the dispersion of density during MD runs). These large error bars are the result of genuine fluctuations of the density of the system due to its relatively small size. Fluctuations must decay with system size as $N^{-1/2}$, hence going from 128 molecules to, say, 1024 molecules will reduce the error bars by a factor of three which, judging from Fig. 13, might be already sufficient.

Due to our current effort on optimization and efficient parallelization of the TBE program, combined with the natural progress in computer performance, we expect the 1024 water molecule simulation to become feasible in a year or even earlier. If we find that the density–temperature dependence is indeed rendered properly, then having a highly transferable TB model with controllable parameters might help to resolve the long standing density maximum dispute.

VIII. METHANOL AND METHANOL–WATER MIXTURES

IX. CONCLUSIONS

1. We have developed a new water model for use in conjuction with the self consistent polarizable ion tight binding method.[5] The model is specifically designed for condensed phase simulations.

2. The model corresponds to the dipole level of the polarizable ion tight binding theory, hence it takes into account point multipoles up to quadrupoles.[9]

3. A novel feature of the new model is that the O–O pair potential includes an attractive $1/r^6$ term. The prefactor is used as a fitting parameter rather than a fixed term found from a theoretical consideration.

4. The density of liquid water at ambient condition is 1.008 g/cm^3 and is larger than density of ice, 0.940 g/cm^3. Hence, ice floats on water in our model.

5. A good agreement with experiment is obtained for RDF shape and peak positions. The evolution of RDF with temperature is correctly reproduced.

6. The dielectric constant and self diffusion coefficient of liquid water are very well reproduced in the temperature interval between 260 K and 350 K. To our knowledge, it is the first successful computation of this scale performed at a quantum mechanical level.

ACKNOWLEDGMENTS

One of us (AYL) would like to thank Dr. Mario G. Del Pópolo for useful comments on molecular dynamics. Queen's University Belfast is acknowledged for the provision of computational resources on a DELL computer cluster. The work is a part of the CASTech catalysis project supported by the EPSRC through grant EP/G012156/1.

Appendix A: Averaged dipoles

In this Appendix we derive the relation to which we refer in Sec. VII B, namely: if

$$\mathbf{p}_{\rm tot} = \mathbf{p}_{\rm pc} + \mathbf{p}_{\rm ind}\,, \tag{A1}$$

where $\mathbf{p}_{\rm pc}$ and $\mathbf{p}_{\rm ind}$ do not correlate, then

$$\bar{p}_{\rm tot} > \bar{p}_{\rm pc}\,, \tag{A2}$$

where $\bar{p}_{\rm tot}$ and $\bar{p}_{\rm pc}$ are the averaged lengths of respective vectors.

1. Definitions

1. By *averaging* we understand finding the average value both with respect to all N molecules in the system and M time frames:

$$\bar{A} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} A_i(t_j) \tag{A3}$$

where A is a quantity in question.

2. Quantities A and B are *uncorrelated* if

$$\overline{AB} = \bar{A}\,\bar{B}\,. \tag{A4}$$

3. As regards the definition of the *averaged length* of a vector, there might be two possibilities:

$$\text{definition (1)}: \quad \bar{v} \equiv \left(\overline{v^2}\right)^{1/2} \tag{A5}$$

$$\text{definition (2)}: \quad \bar{v} \equiv \overline{|\mathbf{v}|}\,, \tag{A6}$$

where $\mathbf{v}$ is some vector quantity. The above two definitions differ in the order in which the averaging and square root operations are applied.

Which definition of the averaged length is used, (1) or (2), is a matter of choice. In our programs, for instance, we employ definition (2). Below we consider both cases.

2. Definition (1)

Multiplying Eq. (A1) by itself and applying the averaging operation (A3), we have

$$\overline{\mathbf{p}_{\rm tot}^2} = \overline{\mathbf{p}_{\rm pc}^2} + \overline{\mathbf{p}_{\rm ind}^2} + 2\overline{(\mathbf{p}_{\rm pc}\cdot\mathbf{p}_{\rm ind})}\,. \qquad \text{(A7)}$$

The averaged scalar product $\mathbf{p}_{\rm pc}\cdot\mathbf{p}_{\rm ind}$ is zero because $\mathbf{p}_{\rm pc}$ and $\mathbf{p}_{\rm ind}$ are uncorrelated. Indeed, using Eq. (A4) we find:

$$\overline{(\mathbf{p}_{\rm pc}\cdot\mathbf{p}_{\rm ind})} = \bar{\mathbf{p}}_{\rm pc}\cdot\bar{\mathbf{p}}_{\rm ind} = 0\,, \qquad \text{(A8)}$$

where the second equality follows from the fact that our system is isotropic, therefore $\bar{\mathbf{p}}_{\rm pc} = \bar{\mathbf{p}}_{\rm ind} = 0$.

Using length definition (A5), we obtain from Eqs. (A7) and (A8):

$$\bar{p}_{\rm tot} \equiv \left(\overline{\mathbf{p}_{\rm tot}^2}\right)^{1/2} = \left(\overline{\mathbf{p}_{\rm pc}^2} + \overline{\mathbf{p}_{\rm ind}^2}\right)^{1/2} > \left(\overline{\mathbf{p}_{\rm pc}^2}\right)^{1/2} \equiv \bar{p}_{\rm pc}$$

as $\mathbf{p}_{\rm ind}$ is not a zero vector.

3. Definition (2)

Here we additionally take advantage of the fact that $|\mathbf{p}_{\rm ind}|/|\mathbf{p}_{\rm pc}| \ll 1$ (see Fig. 9) and shall prove (A2) to the first order in $|\mathbf{p}_{\rm ind}|/|\mathbf{p}_{\rm pc}|$.

Consider a set of vectors $\{(\mathbf{a}_i, \mathbf{b}_i),\, i = 1, \ldots, N\}$ having the property that $b_i/a_i \ll 1$ for any i. Let us create set $\mathbf{c}_i$ as $\mathbf{c}_i = \mathbf{a}_i + \mathbf{b}_i$ and find $\bar{c}$ according to definition (2):

$$\begin{aligned}\bar{c} &= \tfrac{1}{N}\sum_i c_i = \tfrac{1}{N}\left(a_i^2 + b_i^2 + 2\mathbf{a}_i\cdot\mathbf{b}_i\right)^{1/2} \\ &\simeq \tfrac{1}{N}\sum_i \left(a_i + \tfrac{b_i^2}{2a_i} + \tfrac{\mathbf{a}_i}{a_i}\cdot\mathbf{b}_i\right) > \bar{a} + \overline{\mathbf{b}\cdot\mathbf{a}/a}\end{aligned}$$

Thus, $\bar{p}_{\rm tot} > \bar{p}_{\rm pc}$ to the first order in $|\mathbf{p}_{\rm ind}|/|\mathbf{p}_{\rm pc}|$ if vectors $\mathbf{p}_{\rm ind}$ do not correlate with the *direction* of vectors $\mathbf{p}_{\rm pc}$.

REFERENCES

[1] T. J. Sheppard, A. Y. Lozovoi, J. J. Kohanoff, and A. T. Paxton, Universal tight binding model for chemical reactions in solution and at surfaces: I. Organic molecules, (first paper in this series).

[2] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**(24), 6269–6271 (1987).

[3] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**(2), 926–935 (1983).

[4] A. T. Paxton and J. J. Kohanoff, J. Chem. Phys. **134**(4), 044130 (2011).

[5] M. W. Finnis, A. T. Paxton, M. Methfessel, and M. van Schilfgaarde, Phys. Rev. Lett. **81**(23), 5149–5152 (1998).

[6] A. T. Paxton, In *Multiscale Simulation Methods in Molecular Sciences*, eds. J. Grotendorst, N. Attig, S. Blugel, and D. Marx, volume 42 of *NIC series*, pages 145–174, Institute for Advanced Simulation, Forschungszentrum Julich. Available on-line at http://www.fz-juelich.de/nic-series/volume42 (2009).

[7] A. Y. Lozovoi, T. J. Sheppard, J. J. Kohanoff, and A. T. Paxton, Universal tight binding model for chemical reactions in solution and at surfaces: III. Transition metal oxide surfaces, (third paper in this series).

[8] S. Fabris, A. T. Paxton, and M. W. Finnis, Phys. Rev. B **63**(9), 941011–9410113 (2001).

[9] M. Finnis, *Interatomic Forces in Condensed Matter*, Oxford University Press, Oxford, U.K. (2003).

[10] H.-P. Schwefel, *Evolution and optimum seeking*, Wiley (1995).

[11] W. Klopper, J. G. C. M. van Duijneveldt-van de Rijdt, and F. B. van Duijneveldt, Phys. Chem. Chem. Phys. **2**(10), 2227–2234 (2000).

[12] C. J. Burnham and S. S. Xantheas, J. Chem. Phys. **116**(4), 1479–1492 (2002).

[13] L. Goodwin, A. J. Skinner, and D. G. Pettifor, Europhys. Lett. **9**, 701–706 (1989).

[14] D. J. Chadi, Phys. Rev. Lett. **41**(15), 1062–1065 (1978).

[15] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**(15), 154104 (2010).

[16] The cutoff functions used in the present study are fifth order polynomials $P_5(x)$ that produce a smooth step between 1 and 0 with continuous second derivatives:

$$P_5(0) = 1, \qquad P_5'(0) = P_5''(0) = 0,$$
$$P_5(1) = 0, \qquad P_5'(1) = P_5''(1) = 0.$$

Any function $f(r)$ can be smoothly cut to zero between cutoff radii r_1^c and r_c^2 as

$$\tilde{f}(r) = f(r)\, P_5\left[(r - r_c^1)/(r_c^2 - r_c^1)\right].$$

It is easy to demonstrate that function $\tilde{f}(r)$ and its first two derivatives coincide with those of $f(r)$ at $r < r_c^1$, are equal to 0 at $r > r_c^2$, and are continuously differentiable up to the second order in the whole interval where $f''(r)$ exists and is continuous.

[17] D. R. Lide, editor, *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL, 83rd edition (2002).

[18] G. Herzberg, *Molecular Spectra and Molecular Structure II: Infrared and Raman spectra of polyatomic molecules*, Krieger Publishing, Malabar, Florida (1991).

[19] In an *orthogonal* tight binding, the overlap matrix is assumed to be the unit matrix. *Non orthogonal* tight binding has better transferability but computationally is more expensive. Therefore we usually do not use the non orthogonal TB unless unavoidable.

[20] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, Molecular Physics **87**(5), 1117–1157 (1996).

[21] P. Piecuch, S. A. Kucharski, K. Kowalski, and M. Musiał, Comp. Phys. Commun. **149**(2), 71 – 96 (2002).

[22] S. S. Xantheas, C. J. Burnham, and R. J. Harrison, J. Chem. Phys. **116**(4), 1493–1499 (2002).

[23] P. J. Linstrom and W. G. Mallard, editors, *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, National Institute of Standards and Technology, http://webbook.nist.gov/chemistry (2011).

[24] R. M. Olson, J. L. Bentz, R. A. Kendall, M. W. Schmidt, and M. S. Gordon, J. Chem. Theory Comput. **3**(4), 1312–1328 (2007).

[25] B. Santra, A. Michaelides, M. Fuchs, A. Tkatchenko, C. Filippi, and M. Scheffler, J. Chem. Phys. **129**(19), 194111 (2008).

[26] T. K. Hirsch and L. Ojamäe, J. Phys. Chem. B **108**(40), 15856–15864 (2004).

[27] Z. Ma, Y. Zhang, and M. E. Tuckerman, J. Chem. Phys. **137**(4), 044506 (2012).

[28] F. Labat, C. Pouchan, C. Adamo, and G. E. Scuseria, J. Comput. Chem. **32**(10), 2177–2185 (2011).

[29] Z. Raza, D. Alfè, C. G. Salzmann, J. Klimeš, A. Michaelides, and B. Slater, Phys. Chem. Chem. Phys. **13**, 19788–19795 (2011).

[30] C. M. B. Line and R. W. Whitworth, J. Chem. Phys. **104**(24), 10008–10013 (1996).

[31] P. J. Feibelman, Science **295**, 99–102 (2002).

[32] S. M. Jackson, V. M. Nield, R. W. Whitworth, M. Oguro, and C. C. Wilson, J. Phys. Chem. B **101**(32), 6142–6145 (1997).

[33] A. K. Soper and C. J. Benmore, Phys. Rev. Lett. **101**, 065502 (2008).

[34] A. V. Gubskaya and P. G. Kusalik, J. Chem. Phys. **117**, 5920 (2002).

[35] D. P. Fernández, Y. Mulev, A. R. H. Goodwin, and J. M. H. Levelt Sengers, J. Phys. Chem. Ref. Data **24**(1), 33–69 (1995).

[36]K. T. Gillen, D. C. Douglass, and M. J. R. Hoch, J. Chem. Phys. **57**(12), 5117–5119 (1972).
[37]R. Mills, J. Phys. Chem. **77**(5), 685–688 (1973).
[38]P. L. Silvestrelli and M. Parrinello, Phys. Rev. Lett. **82**, 3308–3311 (1999).
[39]M. Sharma, R. Resta, and R. Car, Phys. Rev. Lett. **98**(24) (2007).
[40]H.-S. Lee and M. E. Tuckerman, J. Chem. Phys. **126**(16), 164501 (2007).
[41]M.-V. Fernández-Serra and E. Artacho, J. Chem. Phys. **121**(22), 11136–11144 (2004).
[42]R. Jonchiere, A. P. Seitsonen, G. Ferlat, A. M. Saitta, and R. Vuilleumier, J. Chem. Phys. **135**(15), 154503 (2011).
[43]J. Wang, G. Román-Pérez, J. M. Soler, E. Artacho, and M.-V. Fernández-Serra, J. Chem. Phys. **134**(2), 024516 (2011).
[44]I.-C. Lin, A. P. Seitsonen, M. D. Coutinho-Neto, I. Tavernelli, and U. Rothlisberger, J. Phys. Chem. B **113**(4), 1127–1131 (2009).
[45]A. K. Soper, Chem. Phys. **258**(2-3), 121–137 (2000).
[46]ISIS Disordered Materials Database, www.isis.stfc.ac.uk/groups/disordered-materials/database/.
[47]M. Ji, K. Umemoto, C.-Z. Wang, K.-M. Ho, and R. M. Wentzcovitch, Phys. Rev. B **84**, 220105 (2011).
[48]C. Pinilla, A. H. Irani, N. Seriani, and S. Scandolo, J. Chem. Phys. **136**(11), 114511 (2012).
[49]O. Mishima and H. E. Stanley, Nature (London) **396**, 329–335 (1998).
[50]M. Uematsu and E. U. Frank, J. Phys. Chem. Ref. Data **9**(4), 1291–1306 (1980).
[51]K. R. Harris and L. A. Woolf, J. Chem. Soc., Faraday Trans. 1 **76**, 377–385 (1980).
[52]M. Holz, S. R. Heil, and A. Sacco, Phys. Chem. Chem. Phys. **2**, 4740–4742 (2000).
[53]M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press, New York, U.S.A. (2010).
[54]R. J. Speedy and C. A. Angell, J. Chem. Phys. **65**, 851–858 (1976).